

Sveučilište u Zagrebu
Filozofski fakultet
Odsjek za psihologiju

PSIHOMETRIJSKA VALIDACIJA TESTA S

Diplomski rad

Maša Tonković

Mentor: Doc. dr. Damir Ljubotina

Zagreb, 2007.

SADRŽAJ

SAŽETAK

1. UVOD	1
O inteligenciji	1
O testovima inteligencije	3
O Testu S	6
2. CILJ ISTRAŽIVANJA	7
3. METODOLOGIJA	8
Mjerni instrument – Test S	8
Sudionici i postupak	9
4. REZULTATI I RASPRAVA	11
Analiza Testa S	11
Analiza zadataka	13
Faktorska analiza	17
Skraćenje testa	20
Potencijalna primjena Testa S u praksi	25
5. ZAKLJUČAK	27
6. LITERATURA	28
7. PRILOG	29

Psihometrijska validacija Testa S
The Psychometric Validation of the Test S

Maša Tonković

SAŽETAK

Iskustva psihologa u praksi pokazala su da sadržaj testova s vremenom postaje poznat, stoga postoji stalna potreba za razvojem novih testova inteligencije. Cilj ovog rada je provesti analizu preliminarne forme Testa S, novog testa apstraktnog mišljenja, i njegovog 61 zadatka, kako bi se stvorila konačna, ekonomičnija, verzija testa.

Test je primijenjen na 172 sudionika, učenika viših razreda srednje škole i studenata različitih usmjerenja, u trajanju od sat vremena. Dobivena je normalna distribucija rezultata, a prosječni indeks lakoće ($p=0,675$) govori da test sudionicima nije bio pretjerano težak. Cronbachov koeficijent ukazuje na vrlo visoku pouzdanost testa ($\alpha=0,908$). Osjetljivost testa, prema Fergusonovom koeficijentu također je visoka ($\delta=0,980$). Faktorskom analizom dobiva se veći broj faktora, no veličine njihovih karakterističnih korijena ukazuju na postojanje jednog generalnog faktora, koji objašnjava 17% varijance.

Na osnovi rezultata predložena je skraćena verzije testa od 44 zadatka, predviđenog trajanja primjene 30-35 minuta. Očuvala se normalna distribucija rezultata, a težina se testa povećala prema optimalnoj. Izbacivanjem 17 zadataka pouzdanost i osjetljivost testa su se čak povećale ($\alpha=0,912$; $\delta=0,980$), a u faktorskoj strukturi još više do izražaja dolazi generalni faktor, koji objašnjava 22% varijance. Možemo zaključiti da je skraćivanjem Test S dobio na ekonomičnosti, uz povećanu kvalitetu.

Preporuča se provjeriti rezultate dobivene u ovom radu na reprezentativnom uzorku, kao i funkcioniranje kraće verzije u praksi, te provesti standardizaciju.

Ključne riječi: *inteligencija, test apstraktnog mišljenja, analiza čestica, Test S*

SUMMARY

Experience of psychologists in practice reveal that, since the content of tests often becomes publicly known with time, there is a constant need for new intelligence tests. The aim of this research was to make a complete analysis of the preliminary form of the Test S, a new abstract thinking test, and its 61 items, so that the final, more economical version of the test could be made.

The test was administered to 172 participants - senior high-school students and students of different professions, for the duration of one hour. Results of the test are distributed normally. The average difficulty index ($p=0,675$) shows that the participants didn't find the test to be too difficult. The Cronbach alpha index points to high test reliability ($\alpha=0,908$). Test sensitivity is also high, according to Ferguson's index ($\delta=0,980$). The factor analysis gave more factors, but from the sizes of their eigenvalues it can be concluded that there is only one general factor, which describes 17% of the variance.

A short version of the test is suggested, containing 44 items, with estimated test administration of 30-35 minutes. Normal distribution of results has been preserved and the test difficulty increased towards the optimal one. By deleting 17 items, the reliability and the sensitivity of the test have even increased, and in the factor structure, the general factor is even more emphasized with a variance explanation of 22%. We can conclude that by reducing its size, the Test S became more effective.

Replication of this study on a representative sample and test standardization are recommended. It would be highly useful to check how the short version of the test functions in practice,

Keywords: *intelligence, abstract thinking test, item analysis, Test S*

1. UVOD

O inteligenciji

Inteligencija je, prema *Psihologijskom rječniku* "jedan od važnijih psihologijskih konstrukata koji nije jednoznačno definiran. Najčešće označava svojstvo uspješnog snalaženja jedinke u novim situacijama, u kojima ne pomaže stereotipno nagonsko ponašanje, a niti učenjem stečene navike, vještine i znanja. Za razliku od ovih neplastičnih i reproduktivnih oblika ponašanja koji su korisni samo u nekim prilikama, inteligencija je svojstvo jedinke da pronalazi nove prilagođene reakcije u prilikama bilo koje vrste" (prema Petz i sur., 1992, str. 160). Pojedinci se razlikuju u razvijenom stupnju te sposobnosti, a njihove individualne razlike relativno su stabilne i mogu se mjeriti pomoću odgovarajućih testova inteligencije.

Proučavanju prirode kognitivnog funkcioniranja možemo pristupiti sa strukturalističkog ili s funkcionalnog stajališta (Zarevski, 2002). Teoretičari strukturalističkog stajališta inteligenciji, tj. predstavnici psihometrijskog pristupa, razlikuju se međusobno prema mišljenju postoji li jedinstvena opća kognitivna sposobnost. Engleski psiholog Charles Spearman izložio je 1904. godine (prema Rathus, 2000) stajalište prema kojem sva ponašanja koja smatramo inteligentnima imaju u svojoj osnovi jedan zajednički faktor, koji je nazvao generalnim (g) faktorom inteligencije. Drugim riječima, g faktor se u stanovitoj mjeri upotrebljava kod svih intelektualnih zadataka. No, predvidio je postojanje i specifičnog (s) faktora inteligencije, koji opisuje bilo koju sposobnost koja je jedinstvena za izvršavanje određenog zadatka. Takvi specifični faktori, prema Spearmanu, relativno su nezavisni.

Prema nekima, otkriće generalnog (g) faktora inteligencije i statističke analize povezane s time, najznačajniji su nalazi i doprinosi psihologije (npr. Anderson, 1992, prema Gardner, Kornhaber i Wake, 1999). No postoji i grupa teoretičara koji naglašavaju postojanje grupnih faktora inteligencije. Thurstone (1938) tako pretpostavlja 7 primarnih mentalnih sposobnosti: perceptivni, spacijalni, numerički, verbalni faktori, pamćenje te faktori rječitosti i rezoniranja. Guilfordov model strukture intelekta (1967) razlikuje tri aspekta intelekta (operacije, produkte i sadržaje), čijom kombinacijom se dobiva 120 relativno nezavisnih intelektualnih sposobnosti (Gardner,

Kornhaber i Wake, 1999). Jedna od novijih teorija je Gardnerova teorija višestrukih inteligencija (1983), prema kojoj inteligencija nije samo jedan, jedinstveni konstrukt. Gardner govori o sedam različitih inteligencija koje su međusobno relativno nezavisne: lingvistička, logičko-matematička, spacijalna, glazbena, tjelesno-kinestetička, interpersonalna i intrapersonalna (Gardner, Kornhaber i Wake, 1999).

Hijerarhijski modeli nude rješenje koje prihvaća postojanje grupnih faktora, ali i njegovih užih, podređenih, grupnih faktora. Primjer su Vernonov model intelekta (1956) koji podrazumijeva 4 razine faktora inteligencije, Cattellov hijerarhijski model (1971) s 2 glavna subfaktora, fluidnom i kristaliziranom inteligencijom, te Carrollova «troslojna» teorija iz 1993. (Gardner, Kornhaber i Wake, 1999).

S druge strane, u središtu pažnje funkcionalnog pristupa je način na koji se kognitivne funkcije odvijaju, pa se inteligencija tretira kao (Zarevski, 2000):

- a) sposobnost prilagodbe (pojedinaac se prilagođava okolini i/ili okolinu prilagođava sebi);
- b) kapacitet za učenje (mjeranjem brzine učenja, odnosno složenosti operacija koje pojedinac može naučiti, što implicira da pojedinac ne nasljeđuje inteligenciju, nego sposobnost da uči);
- c) apstraktno mišljenje (naglašava sposobnost manipuliranja idejama i simbolima, korištenje pravila edukacije relacija i korelata, rezoniranje i dedukciju);
- d) sposobnost komunikacije (razumijevanje tuđih emocionalnih stanja, problema i potreba; socijalne vještine; socijalna inteligencija i sl.)

Mišljenje mnogih psihologa jest da inteligencija obuhvaća upravo sposobnost apstraktnog mišljenja, što se odražava i u Termanovoj definiciji: "Pojedinaac je inteligentan razmjerno tome u kojoj je mjeri sposoban apstraktno misliti..." (Terman, 1921, prema Gardner, Kornhaber i Wake, 1999, str. 128). Prema Gardneru, Kornhaberu i Wakeu (1999) apstraktno se mišljenje ne može egzaktno definirati, no može se reći da je ključna sposobnost uviđanja različitih odnosa i struktura, posebice onih koji se ne mogu lako uočiti osjetilima. Bazira se na induktivnom rezoniranju, koje uključuje rezoniranje od specifičnih činjenica ka općem zaključku koji bi mogao objasniti te činjenice (Sternberg, 2005). Induktivno rezoniranje često uključuje procese stvaranja i provjere hipoteza, a zaključke donosimo tako da iz niza specifičnih slučajeva generaliziramo neko opće pravilo. S opažanjem daljnjih pojedinačnih slučajeva možemo

proširiti naše pravilo, ili zaključiti nešto o izuzecima općeg pravila. Prema kognitivnim psiholozima, ljudi koriste induktivno rezoniranje kako bi mogli shvatiti golemu raznolikost u svojoj okolini i predviđati događaje.

O testovima inteligencije

U psihologiji, "test predstavlja standardizirani postupak kojim se izaziva određena aktivnost čija se veličina mjeri i vrednuje radi lakše usporedbe individualnih rezultata i smještanja istih u skupinu rezultata" (Petz i sur., 1992, str. 459). Testovi se u psihologiji mogu koristiti za dijagnostičke svrhe (određivanje veličine određene psihičke karakteristike) ili za prognozu (na osnovi rezultata u testu, predviđanje budućeg uspjeha u određenim aktivnostima). Testovi se uglavnom sastoje od određenog broja čestica, gdje se uspjeh u svakoj zbraja u ukupan krajnji rezultat. Oni nemaju nultu točku jer najmanji rezultat u nekom testu ne znači odsutnost predmeta mjerenja, već samo nemogućnost testa da izmjeri toliko malu zastupljenost iste u individui.

Postoji nekoliko podjela testova: prema prirodi zadatka (testovi maksimalnog učinka ili testovi tipičnog ponašanja), prema načinu primjene testova (individualni ili grupni testovi) i prema relevantnosti vremena (testovi brzine i testovi snage). Također, testovi se mogu podijeliti i s obzirom na predmet mjerenja: testovi sposobnosti (senzornih, psihomotornih ili mentalnih), testovi znanja (provjera efekta učenja) te testovi ličnosti (analitički ili sintetički) (Petz, 1992).

Pod testom inteligencije podrazumijevamo svaki standardizirani postupak za mjerenje individualnih razlika u intelektualnom funkcioniranju (prema Petz, 1992). Iako postoje neslaganja o naravi inteligencije, psiholozi svakodnevno primjenjuju tisuće testova inteligencije (Rathus, 2000). Najčešće se rezultat na testu inteligencije izražava kvocijentom inteligencije (IQ), koji zapravo daje podatak o postotku populacije koji postiže isti ili viši rezultat.

Usredotočivši se na zadatke praktičnog znanja, Alfred Binet i Theodore Simon objavili su 1905. svoje prve testove inteligencije, no i prve testove inteligencije (kako ih mi shvaćamo danas) uopće. Danas se koristi na tisuće različitih testova kognitivnih sposobnosti. Jedan od najpoznatijih i najčešće korištenih testova apstraktnog mišljenja

su Ravenove progresivne matrice (1938). Ravenove progresivne matrice test su koji od ispitanika zahtijeva uviđanje odnosa u apstraktnim vizualnim prikazima. Uključuju mjerenje eduktivne, neverbalne sposobnosti, tj. sposobnosti razumijevanja kompleksnih situacija, sposobnosti pronalaženja značenja u događajima i sposobnosti percepcije i mišljenja. Svi zadaci iz tog testa od ispitanika zahtijevaju analizu, usporedbu i rješavanje analogija koje se temelje na apstraktnim likovima složenim u obliku matrice (Raven, J., Raven, J. C. i Court, 1998).

Istina je da postoji veći broj pouzdanih testova kognitivnih sposobnosti, koji su se u praksi pokazali valjanima i prognostički vrijednima. No, s obzirom da sadržaj testova s vremenom postaje poznat, postoji stalna potreba za novim testovima. Naime, postoje podaci istraživanja koji govore da se vježbom može utjecati na rezultat u rješavanju testova inteligencije na tri načina (Zarevski, 2000):

- a) višekratnim zadavanjem određenog testa ili njegovih paralelnih formi
- b) analizom pogrešaka u rješavanju testova
- c) širokom raspravom o načelima rješavanja određenih mentalnih zadataka, tj. svojevrsnim "treningom inteligencije" ili tzv. vođenjem (engl. *coaching*)

Iskustva psihologa u praksi pokazala su da sadržaj testova s vremenom postaje poznat, a norme za interpretaciju zastarjele, stoga postoji stalna potreba za razvojem novih testova inteligencije. Za dio testova kod nas ne postoje valjane nacionalne norme, kao ni podaci o psihometrijskim karakteristikama. Načela suvremene kvantitativne teorije testova uključuju: konstrukciju preliminarne verzije testa, primjenu na uzorku u svrhu provjere karakteristika, psihometrijsku analizu te izbor zadataka za konačnu verziju testa, primjenu na reprezentativnom uzorku, eventualnu izradu paralelnih verzija, te prilagodbu testa za primjenu putem računala). Upravljanje ljudskim potencijalima, osobito izbor i selekcija kadrova u području gospodarstva, te nekih područja obrazovanja i zdravstva, počiva na procesu psihodijagnostike. Korištenjem adekvatnog psihologijskog instrumentarija taj proces postaje znatno kvalitetniji, brži i djelotvorniji. Testovi koji su težinski prilagođeni potrebama korisnika te standardizirani na reprezentativnim uzorcima omogućuju podizanje kvalitete psihodijagnostike te optimalizaciju procesa donošenja odluka. Mjerenju g-faktora često se pribjegava iz

razloga što nisu poznati specifični kriteriji (npr. uspjeha u studiju / poslu), a pretpostavka je da se generalni faktor inteligencije reflektira u svim specifičnim sposobnostima.

Psihometrijska evaluacija čestica kompozitnih mjernih instrumenata jedan je od središnjih postupaka pri izradi novih ili reviziji postojećih psihologijskih instrumenata (Nunnally i Bernstein, 1994), pa tako i testova inteligencije. Analiza zadataka predstavlja skup postupaka kojima je glavni cilj procjena osobina čestica u svrhu njihove selekcije i komponiranja u cjelovite mjerne postupke, koji će imati neke poželjne karakteristike. Poznavanje osobina pojedinih zadataka važno je zbog činjenice da su sve kvantitativne osobine testovnih rezultata determinirane osobinama njihovih elemenata, tj. pojedinih čestica. Izbor čestica za konačnu formu nekog mjernog instrumenta ovisi o prvom redu o njihovoj osjetljivosti, te diskriminativnoj valjanosti, a posljedice tog izbora odnose se na valjanost, pouzdanost i osjetljivost ukupnih testovnih rezultata. Da bi testovi poslužili svojoj svrsi, oni moraju biti pouzdani, osjetljivi i valjani. Pouzdanost testa odnosi se na stupanj u kojem test daje konzistentne rezultate. Ako je test visoko pouzdan, rezultat se može protumačiti kao mjera uratka pojedinca u testu, a ne slučajnih faktora koji nemaju veze s uratkom pojedinca. Osjetljivost je karakteristika testa ili procedure mjerenja koja se odnosi na sposobnost testa da razlikuje ispitanike prema predmetu mjerenja. Valjanost testa odnosi se na ideju da se testom uistinu mjeri ono što se njime namjerava mjeriti. Ako test zaista mjeri ono što bi trebao mjeriti, na temelju uratka u testu možemo izvoditi zaključke u prilikama izvan testovne situacije. Pouzdanost i valjanost međusobno su povezane. Pouzdanost je nužan preduvjet valjanosti: nemoguće je znati mjeri li test ono za što se tvrdi da mjeri ako su rezultati dobiveni tim testom nepouzdan. No obrnuto ne vrijedi, pouzdanost sama po sebi ne garantira valjanost (Nunnally i Bernstein, 1994).

O Testu S

Test S (Test Slova) je jedan u nizu testova kognitivnih sposobnosti konstruiran na Odsjeku za psihologiju Filozofskog fakulteta u Zagrebu (uz M-seriju (1967), Problemni test (1966, 1971; paralelne forme 1994, 1995), B-seriju (1966), Neverbalnu seriju za ispitivanje inteligencije - NSI (1963) i druge). Autor testa je Damir Ljubotina. Test je konstruiran u sklopu rada Centra za psihodijagnostičke instrumente, te će se njegov daljnji razvoj uklopiti u projekt *Razvoj, standardizacija i psihometrijska validacija testova kognitivnih sposobnosti*. Test, tipa papir-olovka, namjenjen je mjerenju sposobnosti apstraktnog mišljenja, no do sad nije primjenjivan, tako da ne postoje podaci o psihometrijskim karakteristikama čestica, niti samog testa.

Zanimljivost Testa S je u tome što kao podražajni materijal koristi slova. No, njihov poredak u abecedi je irelevantan, pa zapravo ne govorimo o verbalnom testu.

Test S je test na koji ispitanici dobro reaguju, te ima visoku pojavnu i sadržajnu valjanost. No, pokazalo se da je preliminarna forma testa, koja sadrži 61 zadatak, preduga - kako ispitanicima kojima pri kraju testiranja padne produktivnost, tako i psiholozima kojima je neisplativ zbog dugotrajne primjene.

Stoga je svrha ovog rada provesti kompletnu analizu testa i njegovih zadataka, kako bi se stvorila konačna, ekonomičnija, verzija testa na kojoj će biti moguće provesti standardizaciju.

2. CILJ ISTRAŽIVANJA

Glavni je cilj istraživanja provesti kompletnu analizu svih 61 zadataka Testa S, te na temelju toga predložiti skraćenu verziju testa. Također, provjerit će se i konstruktna valjanost testa.

Primjena preliminarne forme Testa S ukazala je na potrebu za kraćom verzijom testa. Često je u praksi test inteligencije samo dio baterije testova, te je previše odvojiti sat vremena samo za njegovu primjenu. Primjena u trajanju od jednog školskog sata (ili nešto kraća, da unutar školskog sata stane i uputa) znatno bi povećala ekonomičnost testa. Naravno, uz uvjet da se očuva kvaliteta testa – u psihometrijskim terminima, pouzdanost, osjetljivost i valjanost.

Analiza čestica Testa S, osim za skraćanje testa, može poslužiti i kao osnova za sastavljanje lakše i teže forme testa, kao i za izradu paralelnih formi.

Saževši sve navedeno, problemi ovog rada su:

- 1) provjeriti psihometrijske karakteristike 61 zadatka preliminarne forme Testa S, što uključuje izračunavanje indeksa lakoće i koeficijenta diskriminativne valjanosti, te empirijskih karakterističnih krivulja zadataka
- 2) provjeriti deskriptivne parametre i psihometrijske karakteristike ukupnih rezultata na Testu S: pouzdanost tipa unutarnje konzistencije, osjetljivost, te konstruktnu valjanost
- 3) na temelju analize zadataka i ukupnih testovnih rezultata sastaviti konačnu verziju Testa S, te provjeriti kako se izbacivanjem zadataka mijenjaju deskriptivni parametri i psihometrijske karakteristike testa

3. METODOLOGIJA

Mjerni instrument – Test S

Testom S ispituje se ispitanikova sposobnost logičkog rezoniranja. U svakom zadatku navedene su skupine slova koje imaju neko zajedničko obilježje. Zadatak sudionika je da to zajedničko obilježje utvrdi i od ponuđenih odgovora s desne strane odabere jedan odgovor koji također posjeduje isto obilježje. Test sadrži 61 zadatak, uz 2 zadatka za primjer koja se ne boduju.

Primjer A:

AA aA aAa AaA a		a) ab	b) aca	c) BB	d) aa	e) fa	f) aac
-----------------	--	-------	--------	-------	-------	-------	--------

U primjeru A, zajedničko obilježje 5 skupina slova s lijeve strane jest da sva sadrže samo slovo «a». Stoga bi od ponuđenih odgovora s desne strane trebalo odabrati odgovor označen slovom d), zato što je to jedini od ponuđenih odgovora koji se također sastoji samo od slova «a».

Primjer B:

a s x c z		a) B	b) D	c) L	d) W	e) o	f) H
-----------	--	------	------	------	------	------	------

U primjeru B, zajedničko obilježje slova s lijeve strane jest da se radi o malim slovima. Stoga bismo od ponuđenih odgovora s desne strane odabrali odgovor označen slovom e), zato što je to jedini od ponuđenih odgovora koji predstavlja malo slovo.

U uputi je napomenuto da ta zajednička obilježja mogu biti vrlo različita, te se mijenjaju tijekom testa.

Da bi se test uspješno riješio, poželjno je, no nije nužno, poznavati abecedu. No, svakako je potrebno znati razlikovati velika od malih slova, te poznavati koje veliko i malo slovo označavaju isto slovo abecede. Stoga test nije namijenjen nepismenima, a za testiranje osoba iz nelatiničnih kultura potrebno će biti izraditi posebne norme.

Sudionicima se u uputi kaže da, zbog ograničenog vremena za rad, rade brzo, ali i točno. Napomene im se da, ako zapnu na nekom zadatku, krenu rješavati nadolazeće zadatke, te se poslije vrate na zadatak koji nisu riješili.

Od sudionika su prikupljeni podaci o njihovom spolu, godini rođenja i školi/fakultetu.

Sudionici i postupak

Test S je preliminarno primijenjen na pet sudionika, tzv. "primjena jedan na jedan". Sudionici su bili dva studenta i troje odraslih ljudi visoke stručne spreme. Uzorak pojedinaca za koje pretpostavljamo da su iznadprosječno inteligentni izabran je s ciljem provjere postoje li eventualna alternativna rješenja zadataka, koje smo prethodno previdjeli. Osim toga, svrha individualne primjene bila je provjeriti je li test razumljiv osobi koja ga rješava, procijeniti težinu testa, te izmjeriti vrijeme potrebno da sudionici dođu do zadnjeg zadatka. Zadatak svakog ispitanika bio je pročitati uputu i riješiti 62 zadatka, te procijeniti za svaki koliko mu je bio težak: lagan, umjereno težak ili težak.

Na temelju rezultata i komentara sudionika iz prve faze testa, zadaci su okvirno poredani po težini od lakših prema težima. U nekoliko zadataka sudionici su uočili alternativne odgovore, pa su u tim zadacima promijenjeni distraktori, a jedan je zadatak u cjelosti izbačen. Uz navedene preinake stvorena je nova, aktualna, verzija Testa S, koja sadrži 61 zadatak.

S obzirom da metrijske karakteristike testa ovise o uzorku na kojem je primijenjen, da bi se mogla provesti analiza čestica testa potreban je heterogeni uzorak. Naime, homogeni uzorak dovodi do matrice interkorelacija specifičnih za taj uzorak, što test može činiti naoko nediskriminativnim.

Stoga uzorak u ovom istraživanju čine učenici i studenti različitih usmjerenja. Od ukupno 172 sudionika, njih 65 su učenici Klasične gimnazije u Zagrebu, 29 učenici Upravno-birotehničke škole (smjer tajnik/tajnica), 25 studenata Fakulteta elektrotehnike i računarstva, te 53 studenta psihologije Filozofskog fakulteta u Zagrebu. Testiranja, u trajanju od sat vremena, provedena su grupno, na matičnim fakultetima / u školama

sudionika. Svi sudionici potpisali su obaviješteni pristanak, u kojem izjavljuju da su svjesni da će se njihovi podaci koristiti u istraživačke svrhe.

Minimalna dob sudionika je 17 godina, što je dob u kojoj je, prema većini teoretičara, sposobnost apstraktnog mišljenja već formirana (Sternberg, 2005). Maksimalna dob je 28 godina, a sudionici su prosječno stari 19 godina i 9 mjeseci. ($M=19,77$, $SD=2,177$; $C=19$; $D=18$). U uzorku ima više sudionica (75,4%), nego sudionika.

4. REZULTATI I RASPRAVA

Analiza Testa S

Podaci prikupljeni Testom S obrađeni su osobnim računalom pomoću programskog paketa SPSS za Windows verzija 12.0.

Ukupni rezultat na Testu S formiran je jednostavnom linearnom kombinacijom 61 zadatka. Deskriptivna statistika tako formiranih ukupnih rezultata prikazana je u Tablici 1.

Tablica 1. Deskriptivna statistika originalne verzije Testa S

	N	min	max	M	SD	prosječni p	prosječna r_{ij}	min r_{ij}	max r_{ij}	α	prosječna r_{iu}
Test S	172	21	61	41,15	9,604	0,675	0,164	-0,173	0,581	0,908	0,366

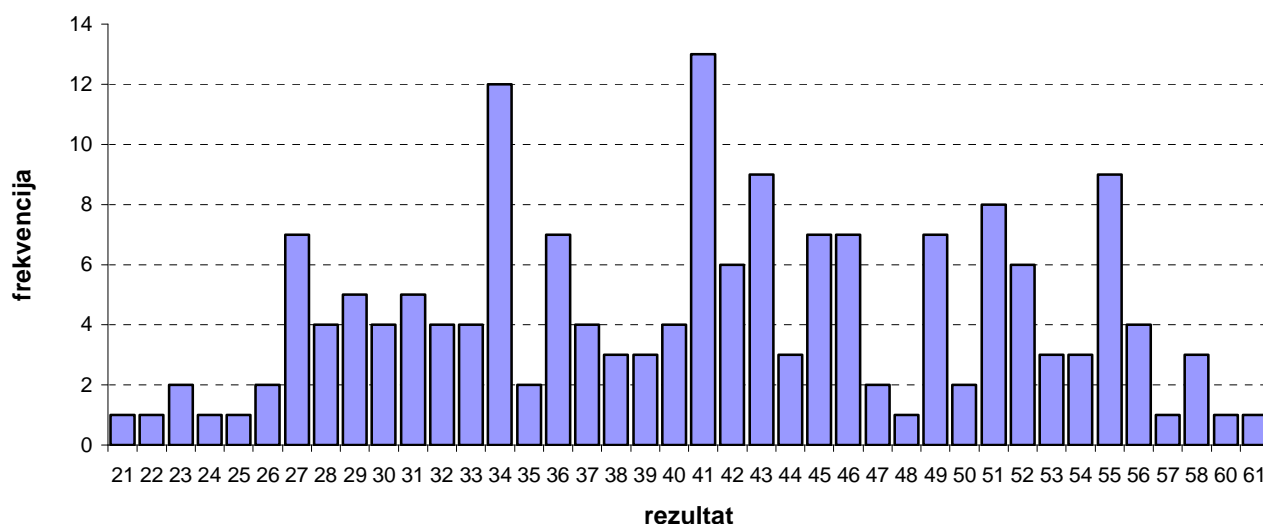
LEGENDA:

N ... broj ispitanika
M ... aritmetička sredina
SD ... standardna devijacija
 α ... Cronbachov Alfa koeficijent pouzdanosti
 r_{ij} ... interkorelacije zadataka
p ... indeks lakoće zadataka
 r_{iu} ... koeficijent diskriminativnosti zadataka

Distribucija ukupnih rezultata na Testu S (Slika 1) statistički se značajno ne razlikuje od normalne (Kolmogorov-Smirnovljeva $Z=1,048$; $p>0,05$). Maksimalni postignuti rezultat iznosi 61, što je ujedno i maksimalni broj bodova koji se uopće može postići na testu. Minimalni rezultat dobiven u testu iznosi 21, a ispitanici su u prosjeku točno riješili 41 zadatak ($M=41,15$; $SD=9,604$). S obzirom da uzorak nije reprezentativan za populaciju, za aritmetičku sredinu dobivenu u ovom istraživanju ne možemo tvrditi da vrijedi i u općoj populaciji, te je ovdje navedena samo radi lakšeg snalaženja u čitanju. Osim toga, rezultati ne mogu biti reprezentativni ni iz razloga što su sudionici rješavali Test S kao test snage, bez vremenskog ograničenja. U praksi će Test S funkcionirati kao test brzine, s ograničenim vremenom rješavanja.

Da su ispitanici test rješavali slučajnim pogađanjem, prosječni indeks lakoće bio bi $p=0,167$, jer je u svakom zadatku ponuđeno 6 odgovora. Ako se primjerena težina testa računa tako da se uzima vrijednost na sredini između indeksa lakoće kojeg bismo dobili da su svi ispitanici sve točno riješili ($p=1$) i slučajnog pogađanja ($p=0,167$)

(Thompson i Levitov, 1985, prema Mathlock-Hetzel, 1997), onda bi idealni prosječni indeks lakoće za Test S iznosio $p=0,583$. Prosječni indeks lakoće dobiven u ovom istraživanju je $p=0,675$ što ukazuje na to da je test lakši nego što bi trebao biti. Ali, mora se uzeti u obzir još jedna implikacija indeksa lakoće (p), a to je da je težina karakteristika i čestice, ali i uzorka na kojem je primijenjen test. Budući da su uzorak činili uglavnom gimnazijalci i studenti, za koje pretpostavljamo da su u prosjeku nadprosječnih kognitivnih sposobnosti, možemo očekivati da su indeksi lakoće viši od onih koje bi dobili primjenom testa na općoj populaciji. Drugim riječima, ako Test S primijenimo na uzorku reprezentativnom za opću populaciju, očekujemo da će se pokazati težim no što se pokazao u ovom istraživanju.



Slika 1. Distribucija rezultata na Testu S (N=172)

Prosječna interkorelacija rezultata iznosi $r_{ij}=0,164$, što upućuje na to da su zadaci unutar testa međusobno u prosjeku nisko povezani, no takav nalaz je očekivan u testovima ovog tipa. Iz tablice 1 vidimo da postoje i zadaci koji su čak u statistički značajnoj negativnoj korelaciji. Najviša korelacija u matrici interkorelacija zadataka testa iznosi $r_{ij}=0,581$. Za usporedbu, prosječna interkorelacija zadataka na Problemnom testu (prema Bosanac, 2006) iznosi $r_{ij}=0,081$ za formu B, te $r_{ij}=0,083$ za formu C. S obzirom da se radi o paralelnim formama Problemnog testa, koji se pokazao pouzdanim i valjanim instrumentom, te se široko primjenjuje u praksi (pogotovo u organizacijskoj psihologiji, u svrhu profesionalne orijentacije i selekcije), možemo zaključiti da su vrijednosti interkorelacija Testa S zadovoljavajuće, pogotovo kad uzmemo u obzir da se

radi o binarnim zadacima koji u pravilu imaju niske prosječne korelacije zbog smanjene varijance.

Pouzdanost je jedna od osnovnih metrijskih karakteristika testa ili mjernog instrumenta općenito, a odnosi se na točnost mjerenja bez obzira na to što se mjeri (Nunnally i Bernstein, 1994). U ovom je radu pouzdanost određena metodom unutarnje konzistencije i izražena je Cronbachovim α koeficijentom. Ova vrsta pouzdanosti dobiva se jednokratnom primjenom upitnika ili testa. Iako u tom slučaju nemamo podatke o vremenskoj stabilnosti testa ili upitnika, imamo o homogenosti, pa se značenje unutarnje konzistencije vjerojatno najviše približava osnovnoj ideji pouzdanosti (Guilford, 1968). Koeficijent unutarnje konzistencije dobiva se na osnovi interkorelacija čestica. Veličina tog koeficijenta ovisi o broju čestica i njihovim korelacijama. Također, pouzdanost će biti veća ako je uzorak ispitanika na kojima se test primjenjuje heterogen. Opće prihvaćeni standardi su da se testovi koji imaju Cronbachov α veći od 0,9 smatraju vrlo visoko pouzdanim, iznad 0,8 visoko pouzdanim, a iznad 0,7 zadovoljavajuće pouzdanim (Bukvić, 1982). Cronbachov koeficijent pouzdanosti dobiven u ovom istraživanju je prilično visok ($\alpha=0,908$), te možemo reći da se radi o vrlo visoko pouzdanom testu, u kojem prosječna korelacija zadatka s ukupnim rezultatom iznosi $r_{iu}=0,366$.

Analiza zadataka

Za svaki zadatak izračunat je indeks lakoće (tj. aritmetička sredina odgovora sudionika na zadatak, što odgovara indeksu lakoće kad su zadaci binarni), te pripadajuća standardna devijacija kao mjera raspršenja rezultata. Uz to, izračunate su i korelacije svakog zadatka s ukupnim rezultatom, tzv. koeficijenti diskriminativne valjanosti zadataka. Navedene vrijednosti nalaze se u Tablici 2.

Najteži zadatak u testu je zadatak broj 60 ($p=0,08$), a najlakši su zadaci broj 9 i broj 13 ($p=1,00$) koje su točno riješili svi ispitanici. Zadatak s indeksom lakoće $p=0$ ili $p=1$ ne doprinosi mjerenju individualnih razlika i zapravo je beskoristan. No, treba se uzeti u obzir da je indeks težine karakteristika i uzorka. Težina zadatka utječe i na varijabilnost testovnih rezultata i na preciznost kojom test razlikuje različite grupe ispitanika. Kad su zadaci u testu ekstremno teški, većina će ukupnih rezultata biti slaba.

I obrnuto, kad su svi zadaci ekstremno lagani, ukupni rezultati će biti visoki. U oba slučaja, testovni rezultati će pokazivati iznimno malo varijabilnosti. Dakle, ekstremne vrijednosti indeksa lakoće direktno sužuju varijabilitet testovnih rezultata (Mathlock-Hetzel, 1997).

Indeksi lakoće viši su u prvoj polovini Testa S, što ne iznenađuje s obzirom da su zadaci intencionalno složeni po težini. No, indeksi težine prema kraju testa ne rastu po redu, pa bi se u idućoj primjeni Testa trebalo razmisliti o tome da se zadaci poredaju prema indeksima težine dobivenim u ovom istraživanju, od lakših prema težima.

Indeksi lakoće umjetno su povećani, jer su ispitanici imali dovoljno vremena za rješavanje. No, u praksi će vrijeme za rad biti ograničeno, pa možemo očekivati manje vrijednosti aritmetičkih sredina svih zadataka. Također, na temelju strogih selekcijskih postupaka za upis na fakultete studenata u uzorku, pretpostavljamo da su sudionici ovog istraživanja u prosjeku nešto viših kognitivnih sposobnosti od opće populacije.

Tablica 2. Prikaz analize 61 zadatka Testa S (N=172)

zadatak	M / p	SD	r_{iu}	α kad se zadatak izbaci	zadržavanje zadatka*	zasićenje s 1. faktorom
1	,9942	,07625	-,095	,909	-	-,117
2	,8837	,32150	,157	,908	-	,180
3	,9070	,29131	,383	,907	+	,431
4	,9942	,07625	,089	,908	-	,101
5	,9535	,21121	,274	,908	+	,303
6	,7151	,45268	,233	,908	-	,257
7	,9826	,13129	,198	,908	-	,212
8	,9767	,15115	,201	,908	-	,198
9	1,0000	,00000	,000	,908	-	,131
10	,9884	,10752	,115	,908	-	,303
11	,8314	,37550	,279	,907	+	,411
12	,7907	,40800	,384	,907	+	,102
13	1,0000	,00000	,000	,908	-	,397
14	,8779	,32835	,093	,909	-	,110
15	,8837	,32150	,363	,907	+	,251
16	,8663	,34135	,116	,909	-	,354
17	,9302	,25550	,237	,908	-	,642
18	,8140	,39028	,327	,907	+	,326
19	,6395	,48154	,603	,904	+	,213
20	,6221	,48628	,305	,907	+	,283
21	,9593	,19817	,199	,908	-	,273
22	,8081	,39491	,258	,908	+	,168
23	,7384	,44080	,254	,908	+	,437
24	,9012	,29931	,181	,908	-	,526

25	,5291	,50061	,391	,906	+	,253
26	,7849	,41210	,486	,906	+	,336
27	,4477	,49871	,231	,908	-	,672
28	,8256	,38058	,319	,907	+	,314
29	,4767	,50092	,625	,904	+	,451
30	,9360	,24538	,286	,907	+	,251
31	,5000	,50146	,404	,906	+	,390
32	,7616	,42733	,231	,908	-	,316
33	,7267	,44693	,365	,907	+	,295
34	,7674	,42370	,285	,907	+	,474
35	,8837	,32150	,257	,908	+	,603
36	,5581	,49806	,438	,906	+	,456
37	,6570	,47611	,564	,905	+	,438
38	,7209	,44985	,420	,906	+	,394
39	,7442	,43759	,400	,906	+	,431
40	,6628	,47414	,350	,907	+	,395
41	,7558	,43086	,394	,906	+	,570
42	,3023	,46061	,363	,907	+	,511
43	,6047	,49035	,517	,905	+	,568
44	,7326	,44392	,456	,906	+	,432
45	,3953	,49035	,524	,905	+	,496
46	,7035	,45805	,390	,906	+	,669
47	,7093	,45541	,458	,906	+	,574
48	,4302	,49655	,612	,904	+	,588
49	,5814	,49477	,525	,905	+	,597
50	,4709	,50061	,536	,905	+	,562
51	,5058	,50143	,548	,905	+	,465
52	,4767	,50092	,512	,905	+	,221
53	,2791	,44985	,417	,906	+	,447
54	,3895	,48907	,193	,909	-	,465
55	,3372	,47414	,408	,906	+	,640
56	,1628	,37025	,419	,906	+	,510
57	,3663	,48319	,598	,904	+	,437
58	,2209	,41608	,476	,906	+	,297
59	,4419	,49806	,392	,906	+	,314
60	,0814	,27424	,264	,908	+	-,117
61	,1628	,37025	,283	,907	+	,180

LEGENDA:

- N ... broj ispitanika
- M ... aritmetička sredina
- p ... indeks lakoće zadataka
- SD ... standardna devijacija
- r_{iu} ... koeficijent diskriminativnosti zadatka
- α ... Cronbachov Alfa koeficijent pouzdanosti
- * ... kriterij zadržavanja zadataka je $r_{iu} > 0,250$
- + ... zadatak ulazi u konačnu verziju testa
- ... zadatak ne ulazi u konačnu verziju testa

Diskriminativna valjanost čestice najčešće se određuje kao korelacija između uratka na pojedinoj čestici i ukupnog rezultata bez te čestice. Visoka pouzdanost (spuriozna korelacija) nam govori u prilog konstruktne valjanosti skupa podataka koji ukazuju na to da li neki mjerni instrument mjeri željeni konstrukt, te u prilog homogenosti instrumenta. Interpretacija koeficijenta korelacije ovisi o vrsti varijabli između kojih se računa korelacija, no katkada se kao gruba orijentacija navodi da koeficijent korelacije od 0,0 do 0,2 ukazuje na nikakvu ili vrlo slabu povezanost, od 0,2 od 0,4 na slabu povezanost, od 0,4 do 0,7 na srednju pouzdanost, a vrijednosti iznad 0,7 na veliku pouzdanost među varijablama (Petz i sur., 1992).

Indeksi diskriminativne valjanosti na Testu S kreću se od $r_{1,u}=(-0,095)$ do $r_{29,u}=0,625$, no u prosjeku su vrlo visoki ($r=0,366$). Najviši koeficijent diskriminativne valjanosti ima 29. zadatak, te je primjer najbolje postavljenog zadatka, tj. onoga koji najbolje razlikuje uspješne od neuspješnih. Ako zadatak mjeri isto što i sam test, onda od ispitanika koji dobro riješe test, očekujemo da će i taj zadatak točno riješiti. Dobar zadatak razlikuje one koji test rješavaju dobro, od onih koji ga rješavaju loše.

Određivanje relacije između pojedinih zadataka i mjerenog konstrukta (ono što zadaci u osnovi tendiraju mjeriti), jedan je od ključnih problema pri konstrukciji nekog mjernog instrumenta. Stoga je jedna od dodatnih metoda provjere valjanosti zadatka inspekcija empirijskih karakterističnih krivulja zadataka. Niti jedan konstrukt (u ovom slučaju inteligencija) ne reflektira se savršeno u nekom konačnom broju zadataka. Laički rečeno, odnos bodova na nekom testu nikad u potpunosti ne odgovara odnosu stvarnoj veličini konstrukta.

Skoro svi modeli skaliranja ispitanika mogu se opisati različitim tipovima krivulja koje povezuju mjereni atribut i vjerojatnost indikativnog (točnog) odgovora. Funkcije ovog oblika nazivaju se linije traga zadataka ili karakteristične krivulje zadataka (Nunnally i Bernstein, 1994). Apscisa se, kod ovakvih prikaza, odnosi na mjereni konstrukt, definiran u terminima njegove veličine (izraženosti), u ovom istraživanju ukupni rezultat na Testu S, kategoriziran prema veličini u 6 razreda. Na ordinati se nalazi vjerojatnost točnog odgovora, konkretno indeks lakoće zadatka. Tako iz krivulje možemo vidjeti koja je vjerojatnost da će osobe određenog stupnja razvijenosti apstraktnog mišljenja točno riješiti zadatak, te kakav je odnos težine

određenog zadatka i veličine ukupnog rezultata. Generalno se očekuje da kako raste mjereni konstrukt (inteligencija), tako raste i vjerojatnost točnog odgovora u zadatku.

Različiti modeli mjerenja dovode do različitih linija traga. Moderni modeli teorije testova pretpostavljaju sigmoidnu krivulju, koja odgovara kumulativnoj krivulji normalne distribucije. Takva linija traga znatno je diskriminativnija (ima veći nagib) u svom središnjem dijelu, nego na krajevima. Strmija linija znači i veću korelaciju zadatka i ukupnog uratka.

Pri odluci koje zadatke u testu zadržati, dodana vrijednost karakterističnih krivulja zadataka jest kojem je uzorku zadatak primjeren – razlikuje li zadatak bolje iznadprosječne (strmiji nagib kod sudionika uspješnih na cijelom testu) ili ispodprosječne sudionike (krivulja u početku ima strmi nagib, no vrlo brzo dostiže plato). Stoga daju vrijednu informaciju pri konstrukciji lakše i teže verzije nekog testa.

Karakteristične krivulje zadataka u Testu S (u prilogu) pokazuju adekvatnost zadataka odabranih prema kriteriju indeksa diskriminativne valjanosti. Krivulje zadržanih zadataka relativno su strme, a kao najbolji zadatak (onaj čija krivulja najviše sliči sigmoidnoj, a proteže se duž cijelog raspona ordinate) pokazao se opet zadatak broj 29.

Faktorska analiza

Da bi se provjerila latentna struktura skupa varijabli (zadataka) te zajednički faktori u pozadini njihove eventualne povezanosti, primjerena obrada rezultata je faktorska analiza. Faktorska analiza predstavlja skup statističko-matematičkih postupaka koji omogućavaju da se, u većem broju međusobno povezanih varijabli, utvrdi manji broj temeljnih varijabli (Fulgosi 1979). Te temeljne varijable objašnjavaju povezanost između opaženih (manifestnih) varijabli i nazivaju se latentne varijable, tj. faktori. Smisao faktorske analize temelji se na postojanju zajedničkih faktora, koji određuju povezanost između određenog broja manifestnih varijabli, tj. objašnjavaju određeni dio totalne varijance tih varijabli. Dakle, cilj nam je provjeriti faktorsku strukturu Testa S, tj. faktorsku valjanost. Kao osnova za faktorsku analizu poslužile su međusobne korelacije 59 zadataka. Zadaci pod brojevima 9 i 13 nisu ušli u daljnju analizu, jer njihova varijanca iznosi 0.

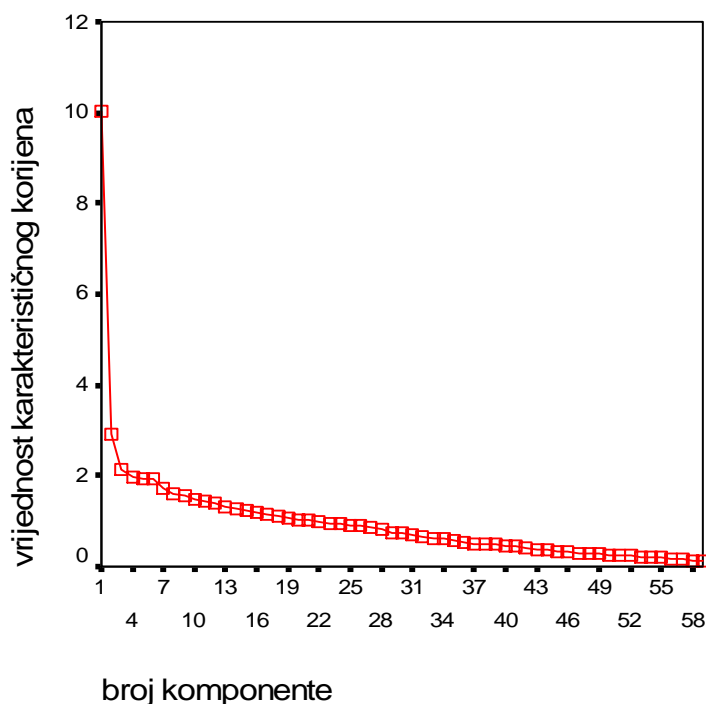
Iako ispitanika ima tek 3 puta više od broja zadataka, Bartlettov test ($\chi^2=3593,705$; $df=1711$; $p<0,01$) i Kaiser-Meyer-Olkinov indeks ($KMO=0,715$) pokazuju da je korelacijska matrica pogodna za provedbu faktorske analize. Još jedan argument u prilog korištenju faktorske analize na ovim podacima je da se ukupni rezultati na testu distribuiraju normalno (Kolmogorov-Smirnovljev $Z=1,048$; $p>0,05$).

Provedena je faktorska analiza testa pod modelom glavnih komponenata, te su dobiveni rezultati prikazani u Tablici 3. Uz Kaiser-Guttmanov kriterij dobiven je 21 faktor. No, veličine njihovih karakterističnih korijena ukazuju na jedan faktor, pošto se vrijednosti korijena ostalih faktora znatno ne razlikuju. Iako prvi faktor ne objašnjava veliki dio ukupne varijance (17%), on brojčano iskače među svim ostalim faktorima. Zasićenja zadataka s glavnim faktorom nalaze se u zadnjem stupcu tablice 2. Očekivano, vidimo da niža zasićenja s 1. faktorom imaju manje diskriminativni zadaci.

Tablica 3: Ekstrahirani faktori na Testu S i njihovi karakteristični korijeni pod modelom glavnih komponenata

komponenta	karakteristični korijeni	% ukupne varijance objašnjen pojedinim faktorom
1	10,029	16,999
2	2,924	4,957
3	2,135	3,619
4	1,963	3,327
5	1,931	3,272
6	1,915	3,246
7	1,734	2,939
8	1,595	2,704
9	1,549	2,626
10	1,493	2,530
11	1,450	2,458
12	1,409	2,388
13	1,323	2,243
14	1,265	2,144
15	1,223	2,072
16	1,184	2,007
17	1,138	1,929
18	1,106	1,875
19	1,084	1,838
20	1,035	1,755
21	1,014	1,718

Sljedeći često korišten kriterij za donošenje odluke o zadržavanju faktora jest izgled tzv. *screeplot* prikaza. Radi se o grafičkom prikazu na čijoj je apscisi redni broj ekstrahiranog faktora, a na ordinati veličina pripadajućeg karakterističnog korijena. Broj faktora koji će se zadržati prekida se kada se karakteristični korijeni prestanu razlikovati i njihov pad postane linearan. Inspekcijom *screeplot* dijagrama na slici 2 primjećujemo da, iako postoji ogroman pad već nakon prvog ekstrahiranog faktora, on postaje linearan nakon ekstrakcije drugog faktora, što ukazuje na to da treba zadržati i drugi faktor.



Slika 2. Grafički prikaz vrijednosti karakterističnih korijena (*screeplot*) za faktore u Testu S

No, treba naglasiti da se u Testu S radi o binarnim zadacima, što znači da je varijanca u zadacima smanjena, te da su njihove međusobne korelacije niže nego što bi bile da su zadaci na nekoj metrički superiornijoj skali. To dovodi do hiperfaktorizacije, tj. do pojavljivanja većeg broja faktora s karakterističnim korijenom iznad 1 (Fulgosi, 1979).

Na temelju svega navedenog, zaključila bih da, iako karakteristični korijeni i *screeplot* dijagram ukazuju na mogućnost postojanja višefaktorskog rješenja, možemo reći da zadaci Testa S u svojoj pozadini zapravo imaju jedan generalni faktor, koji

bismo mogli interpretirati kao g-faktor inteligencije. Moguće je da ispod g-faktora postoje grupni faktori, što bi objasnilo postojanje tako velikog broja faktora sa značajnim karakterističnim korijenima. No, nije ih moguće smisleno interpretirati, pa bi faktorsku strukturu trebalo još jednom provjeriti na većem i reprezentativnom uzorku. Također, valja još jednom provjeriti faktorsku strukturu nakon izbacivanja nediskriminativnih zadataka.

Skraćenje testa

Kao temelj odabira zadataka koje bi se trebalo izbaciti prvenstveno služe koeficijenti diskriminativne valjanosti. Po definiciji taj koeficijent prikazuje koliko je određeni zadatak valjan da razlikuje uspješne od neuspješnih pojedinaca na testu (Bosanac, 2006). Visoki koeficijent znači da zadatak izvrsno diskriminira sudionike s obzirom na njihovu sposobnost koja se mjeri testom. Kao još jedan pokazatelj valjanosti zadatka može poslužiti i podatak o promjeni u Cronbachovom koeficijentu pouzdanosti ako se određeni zadatak izbacuje (Tablica 2).

Tablica 4. Deskriptivna statistika Testa S skraćenog na 44 zadatka (N=172)

	N	min	max	M	SD	prosječni p	prosječna r_{ij}	min r_{ij}	max r_{ij}	α	prosječna r_{iu}
Test S	172	8	44	26,48	8,792	0,602	0,186	-0,132	0,581	0,912	0,412

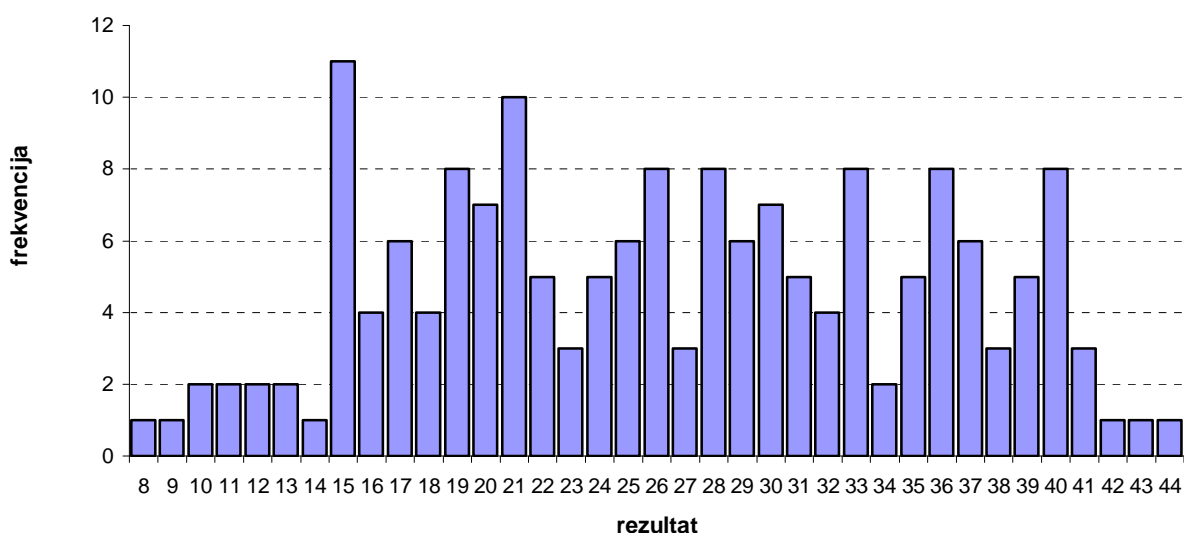
LEGENDA:

N ... broj ispitanika
M ... aritmetička sredina
SD ... standardna devijacija
 α ... Cronbachov Alfa koeficijent pouzdanosti
 r_{ij} ... interkorelacije zadataka
p ... indeks lakoće zadataka
 r_{iu} ... koeficijent diskriminativnosti zadataka

Kao provizoran kriterij zadržavanja zadatka postavljen je uvjet da koeficijent diskriminativne valjanosti bude minimalno $r_{iu}=0,25$. Prema tom, umjereno strogom kriteriju, u konačnu verziju Testa S prolaze 44 čestice. Skraćivanjem su zahvaćeni zadaci svih težinskih razreda, no redukcija se najviše odrazila na lakše zadatke, iz prve trećine testa, jer su se oni pokazali nediskriminativnima. Zadaci srednje težine imaju

veću varijancu od teških i laganih zadataka, stoga je i očekivano da će upravo oni najviše korelirati s ukupnim rezultatom.

Deskriptivna statistika novoformiranog testa prikazana je u Tablici 4. Aritmetička se sredina testa skraćivanjem naravno smanjuje, no veličina smanjenja je povezana i sa smanjenjem prosječnog indeksa lakoće. To je rezultat činjenice da su većinu izbačenih zadataka činili oni s većim indeksom lakoće. Prosječni indeks lakoće iznosi $p=0,602$, što je vrijednost koja se približila idealnoj težini testa. Raspon rezultata smanjuje se skraćanjem testa, kao i standardna devijacija. Oblik distribucije (slika 3) ne mijenja se bitno, skraćivanjem testa zadržali smo normalnu distribuciju ukupnog rezultata na testu (Kolmogorov-Smirnovljev $Z=1,157$; $p>0,05$).



Slika 3. Distribucija rezultata na skraćenoj verziji Testa S od 44 zadatka (N=172)

Prosječna interkorelacija izbacivanjem zadataka raste sa vrijednosti $r_{iu}=0,366$ na $r_{iu}=0,412$, što ukazuje na povećanje dijagnostičke valjanosti testova te njihovo faktorsko čišćenje. Iako su broj zadataka i pouzdanost proporcionalne veličine, izbacivanjem čak 17 čestica pouzdanost Testa S se povećala na $\alpha=0,912$. Cilj psihometrije upravo jest protumačiti i predvidjeti velik broj zbivanja i događaja na osnovi malog broja generalnih varijabli ili uzroka, tj. obuhvatiti što više varijance neke pojave sa što manje zadataka. Standardna pogreška mjerenja iznosi $\sigma_e=2,608$. Ta veličina odgovara očekivanoj standardnoj devijaciji rezultata jedne osobe na velikom broju paralelnih testova (Nunnally i Bernstein, 1994). Iz podatka o veličini standardne pogreške

mjerjenja možemo zaključiti da će se, uz 5% rizika, pravi rezultat sudionika kretati u rasponu od $x \pm 5,11$ bodova, gdje je x ukupni rezultat sudionika u ovoj primjeni Testa S.

Osjetljivost kao mjerna karakteristika definirana je kao mogućnost razlikovanja ispitanika na osnovi njihova ukupnog individualnog rezultata u testu (Nunnally i Bernstein, 1994). Stupanj primjerenosti testa može se prepoznati već po obliku distribucije rezultata i po mjerama raspršenja, standardnoj devijaciji i totalnom rasponu. Prema tim kriterijima, skraćivanjem se osjetljivost Testa S nije bitno promijenila. No, dodatne informacije nam mogu dati broj ostvarenih razlikovanja (BOR) i Fergusonov delta-koeficijent. Navedeni parametri izračunati su za obje verzije testa: originalnu i skraćenu (Tablica 5).

Tablica 5. Podaci vezani uz osjetljivost originalne i skraćene verzije Testa S (N=172)

verzija	min	max	TR	BRR	BOR	δ
originalna (k=61)	21	61	41	40	14256	0,980
skraćena (k=44)	8	44	37	37	14254	0,985

LEGENDA:

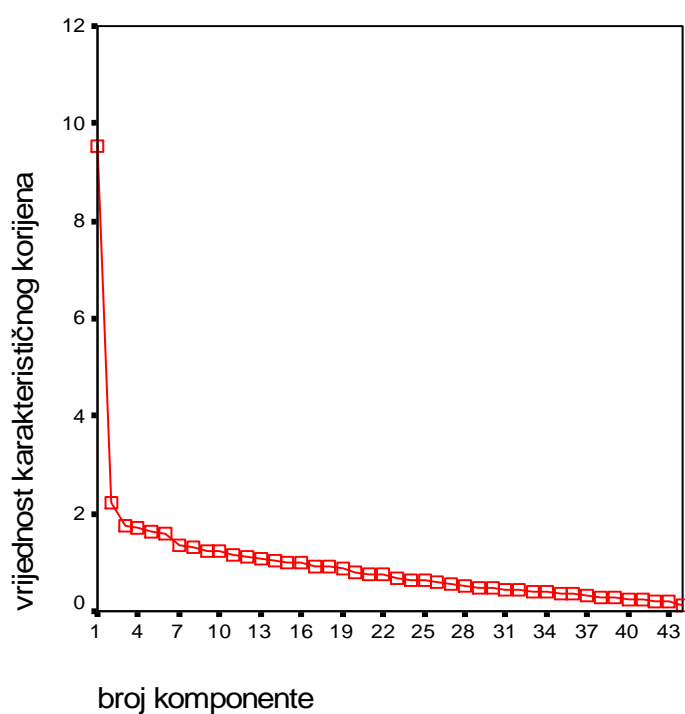
k... broj zadataka u testu
min ... najmanji dobiveni rezultat
max ... najveći postignuti rezultat
TR ... totalni raspon
BRR ... broj različitih rezultata
BOR ... broj ostvarenih razlikovanja
 δ ... Fergusonov koeficijent osjetljivosti

Iako je u novoj verziji testa broj zadataka smanjen, BOR je gotovo identičan u obje verzije testa. Fergusonov koeficijent osjetljivosti, koji u svom računu neutralizira utjecaj smanjenja broja zadataka, skraćivanjem se čak povećao. Takav nalaz može se objasniti činjenicom da su reducirane varijable s ekstremnim indeksima lakoće, koje nisu bile osjetljive. Sve u svemu, može se zaključiti da smo izbacivanjem određenih zadataka dobili čak i na osjetljivosti testa.

Da bismo vidjeli dovodi li skraćivanje testa do nešto jasnije faktorske strukture, ponovno je provedena faktorska analiza, na zadržanih 44 zadataka. U toj, skraćenoj verziji testa, broj ispitanika je 4 puta veći od broja varijabli, pa postoje uvjeti za provođenje faktorske analize. To potvrđuju i $KMO=0,798$ i Bartlettov test sferičnosti ($\chi^2=2531,896$; $df=946$; $p<0,01$).

Tablica 6. Ekstrahirani faktori na skraćenoj verziji Testa S i njihovi karakteristični korijeni pod modelom glavnih komponentata

komponenta	karakteristični korijeni	% ukupne varijance objašnjen pojedinim faktorom
1	9,548	21,701
2	2,229	5,065
3	1,762	4,005
4	1,714	3,895
5	1,642	3,731
6	1,605	3,648
7	1,348	3,063
8	1,320	3,000
9	1,246	2,833
10	1,214	2,758
11	1,171	2,662
12	1,126	2,559
13	1,082	2,460
14	1,051	2,388
15	1,007	2,289



Slika 4. Grafički prikaz vrijednosti karakterističnih korijena (*screeplot*) za faktore u skraćenoj verziji Testa S

Ukupan broj faktora s karakterističnim korijenom većim od 1 se smanjuje sa 21 na 15 (Tablica 6), a generalni faktor dolazi još više do izražaja (Slika 4), objašnjavajući

22% ukupne varijance. Zasićenja gotovo svih varijabli veća su na 1. faktoru, u odnosu na bilo koji idući

Uzevši u obzir ista ograničenja binarnih zadataka kao i kod faktorske analize provedene na svim originalnim zadacima, zaključujemo o postojanju jednog generalnog faktora, pa se može reći da rezultati faktorske analize skraćene verzije Testa S govore u prilog interpretaciji o g-faktoru inteligencije

Kod skraćivanja testa javlja se pitanje što napraviti s prvih nekoliko zadataka, kojima je indeks lakoće maksimalan, a koeficijent diskriminativne valjanosti minimalan (nekima čak i negativan) (Bosanac, 2006). Takvi lagani zadaci imaju ulogu davanja ispitaniku osjećaja uspjeha na početku rješavanja, što facilitira daljnje rješavanje. Također, daju uvid ispitivaču do koje mjere je ispitanik u stanju rješavati zadatke. No s druge strane, ispitanik se često umori rješavajući velik broj zadataka, te zbog lakših zadataka ne stigne riješiti teže. Osim toga, takvi nediskriminativni zadaci narušavaju pouzdanost cijelog testa. Praksa je da se takvi zadaci zadrže u testu, ali se ne boduju.

Iduće pitanje koje nameće skraćivanje testa jest koliko bi trebala trajati njegova primjena. U ovom istraživanju je sat vremena gotovo svim ispitanicima bilo dovoljno za dovršavanje originalne verzije od 61 pitanja. Primjena Testa S kao testa snage bila je nužna radi provedbe analize čestica. No, Test S je planiran primarno kao test brzine, pa bi za primjenu njegove verzije od 44 zadataka trebalo biti dovoljno 30-35 min. Za usporedbu, primjena Problemnog testa, koji sadrži 70 čestica, traje 40 min. No, da bi se utvrdilo vrijeme primjene Testa S koje će dovesti do optimalnog razlikovanja ispitanika, svakako je potrebna provjera u praksi.

Iako istraživanja pokazuju da ne postoje razlike u općoj inteligenciji između muškaraca i žena, poželjno je provjeriti dobivene rezultate na reprezentativnom uzorku. Naime, moguće je da prosjeci na testovima inteligencije budu isti, a da to još uvijek ne znači da je latentna struktura tih sposobnosti ista, što može rezultirati različitim sklopom interkorelacija na testu inteligencije u muškaraca i žena (Zarevski, 2000).

Također, preporuča se provedba faktorske analize na većem broju ispitanika, kako bi se utvrdila opravdanost zaključka o postojanju generalnog faktora. Bolji način provjere valjanosti testa bila bi provjera kongruentne valjanosti, tj. ispitivanje povezanosti rezultata s rezultatima nekog drugog provjerenog testa. S tim ciljem, dio

ispitanika u uzorku bio je zamoljen da uz ostale socio-demografske podatke navede i šifru. Radi se, naime, o studentima 2. godine psihologije koji su ispitivani u okviru kolegija *Uvod u teoriju testova*, na kojem će rješavati još testova inteligencije, koji se onda mogu korelirati s Testom S.

Za buduće primjene Testa S predlaže se zadržavanje nekoliko lakših zadataka (npr. zadaci broj 9, 13 i 17) koji ne bi ulazili u konačni rezultat, te formiranje novog redoslijeda prema indeksima lakoće dobivenim u ovom istraživanju, od viših prema nižima.

Ovisno o namjeri ispitivanja, tj. želimo li bolje razlikovati nadprosječne ili ispodprosječne ispitanike, vrijeme se može skraćivati, odnosno produljivati na određenim populacijama, što se treba utvrditi dodatnim istraživanjima. Također, na temelju provedene analize zadataka, moguće je napraviti lakšu i težu formu testa, kao i paralelne forme.

Jedan od nadolazećih koraka je svakako i standardizacija (kraće verzije) Testa S, s obzirom na spol, dob, mjesto stanovanja, stupanj obrazovanja i ostale demografske varijable. Također, preporuča se adaptacija testa za kompjutersku primjenu, što otvaraju korisnicima nove tehnološke mogućnosti u procesu dijagnostike (mjerenje brzine procesiranja, standardizacija postupka mjerenja, automatizacija procesa obrade i interpretacije rezultata, skraćanje procesa dijagnostike, testiranje na daljinu i dr.).

Potencijalna primjena Testa S u praksi

Ovakav test g-faktora uklapa se u baterije testova gotovo svih psihologa koji provode psihologijsko testiranje. Gotovo da nema selekcijskog postupka koji ne uključuje testiranje inteligencije, najčešće g-faktora. Važne karakteristike testova inteligencije u selekciji su nepoznatost sadržaja i ekonomičnost. Osoba koja traži posao javlja se na nekoliko natječaja i prolazi kroz nekoliko testiranja. U slučaju da na svakom testiranju osoba rješava isti test inteligencije, moguće je da će svaki put postizati sve bolji rezultat, što zbog faktora uvježbavanja, što zbog prorađivanja zapamćenih zadataka u međuvremenu ili razmjene iskustava s prijateljima. Na zadnjem testiranju osoba može postići značajno viši rezultat od svog originalnog, što nije pošteno u odnosu na ostale kandidate koji su na testiranju tek prvi put, a doprinosi i lošijoj prognostičkoj

valjanosti testa. Stoga je važno u selekciju uvoditi nove instrumente. Skraćena verzija Testa S može se koristiti u relativno kratkom vremenu, gdje test funkcionira kao test brzine. Takvom primjenom lako će se prepoznati nadprosječni pojedinci, što često i jest svrha testiranja inteligencije u selekciji.

U kliničkoj psihologiji, Test S bi se mogao koristiti u slučajevima kad treba provjeriti je li klijent iznadprosječno inteligentan. Za to se u praksi široko koriste Ravenove progresivne matrice (RPM), odnosno Progresivne matrice za nadarene (APM), test koji od ispitanika zahtijeva uviđanje odnosa u apstraktnim vizualnim prikazima. Obzirom da su zadaci u testu isključivo apstraktnog oblika, smatra se da je sam test manje ovisan o formalnoj naobrazbi te kulturalnim odrednicama (Franić, 2003), što bi mogla biti odlika i Testa S. Također, s obzirom da je Test S test apstraktnog mišljenja, koji zahtijeva i sposobnost promjene kognitivnih strategija (jer zajednička obilježja slova mogu biti različita i mijenjaju se tijekom testa), pad rezultata može ukazivati na lezije čeonog režnja, kao i na shizofreniju (Galić, 2007).

U školskoj se psihologiji Test S može koristiti za identifikaciju nadarenih učenika, pa tako i kao dio prijemnog ispita za neke fakultete, npr. studij psihologije. Također, može se koristiti i za utvrđivanje stupnja kognitivnog razvoja djeteta. Prema Piagetu (Vasta, Haith, Miller, 1998), s početkom adolescencije, između 12. i 13. godine, počinje razdoblje formalnih operacija. Do tad je dijete ograničeno na bavljenje onim što se nalazi izravno pred njim – s konkretnim. U fazi formalnih operacija adolescent ima sposobnost hipotetičko-deduktivnog rasuđivanja, te se može očekivati porast na testovima apstraktnog mišljenja kakav je Test S, no ne smijemo zaboraviti da se inteligencija razvija do kasnije adolescencije, te je IQ relativno stabilan u dobi između 20 i 60 godina (Zarevski, 1998). Test S bi mogao koristiti školskim psiholozima i pri ispitivanju razloga neuspjeha u srednjoj školi.

Na temelju ovog rada moguće je izraditi lakšu i težu verziju, te paralelne forme Testa S. Sve u svemu, rezultati ukazuju da će Test S, kao pouzdan, osjetljiv, valjan, ali i ekonomičan test novog sadržaja, naći svoju primjenu u organizacijskoj, kliničkoj i školskoj psihologiji.

5. ZAKLJUČAK

Provedena je kompletna analiza 61 zadatka preliminarne verzije Testa S, novog testa apstraktnog rezoniranja, primijenjenog na 172 ispitanika. Distribucija ukupnih rezultata statistički se značajno ne razlikuje od normalne (Kolmogorov-Smirnovljev $Z=1,048$; $p>0,05$). Cronbachov koeficijent pouzdanosti dobiven u ovom istraživanju je prilično visok ($\alpha=0,908$), te možemo reći da se radi o vrlo visoko pouzdanom testu. Osjetljivost testa, prema Fergusonovom koeficijentu također je visoka ($\delta=0,980$). Prosječna korelacija zadatka s ukupnim rezultatom iznosi $r_{iu}=0,366$. Prosječni indeks lakoće iznosi $p=0,675$, što govori da test ispitanicima nije bio pretjerano težak. No, radi se o ispitanicima više inteligencije od prosjeka populacije.

Faktorskom analizom, uz Kaiser-Guttmanov kriterij, dobiva se veći broj faktora. No, veličine njihovih karakterističnih korijena ukazuju na jedan generalni faktor, koji opisuje 17% varijance, te ostale koji se ne razlikuju puno po svojim navedenim vrijednostima.

U ovom je radu na temelju analize čestica donesen prijedlog za formiranje skraćene verzije testa od 44 zadatka, predviđenog trajanja primjene 30-35 minuta. Očuvala sa normalna distribucija rezultata, a težina se testa povećala prema idealnoj. Izbacivanjem 17 zadataka pouzdanost i osjetljivost testa su se čak povećale ($\alpha=0,912$; $\delta=0,085$), a u faktorskoj strukturi još više do izražaja dolazi generalni faktor, koji objašnjava 22% varijance, pa možemo zaključiti da je skraćivanjem Test S dobio na ekonomičnosti, uz povećanu kvalitetu.

Preporučuje se provjeriti rezultate dobivene u ovom radu na reprezentativnom uzorku, kao i funkcioniranje kraće verzije u praksi, te provesti standardizaciju također na reprezentativnom uzorku.

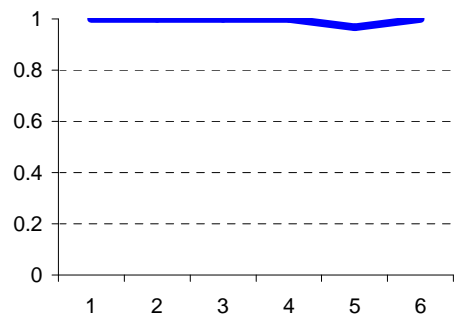
Sve u svemu, Test S je instrument čije karakteristike obećavaju, i trebao bi naći svoje mjesto u bateriji testova organizacijskih, kliničkih i školskih psihologa.

6. LITERATURA

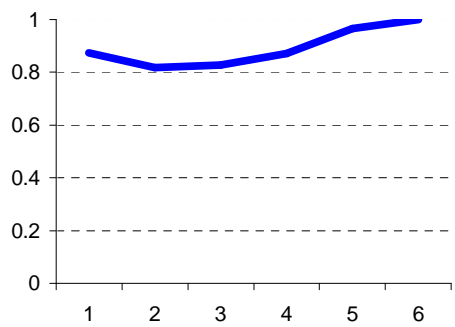
- Bosanac, I. (2006). *Analiza zadataka dviju paralelnih formi Problemnog testa*. Diplomski rad. Filozofski fakultet, Zagreb.
- Bukvić, A. (1982). *Načela izrade psiholoških testova*. Beograd: Zavod za udžbenike i nastavna sredstva.
- Franić, M. (2003). *Provjera diferencijalno-dijagnostičke valjanosti Revidirane Beta Serije na različitim dijagnostičkim skupinama*. Diplomski rad. Filozofski fakultet, Zagreb.
- Fulgosi, A. (1979). *Faktorska analiza*. Zagreb: Školska knjiga.
- Galić, S. (2007). Poremećaji konceptualnog mišljenja i izvršnih funkcija u shizofrenih osoba. *Suvremena psihologija*, 1, 77-94.
- Gardner, H., Kornhaber, M. L., Wake, W. K. (1999). *Inteligencija: različita gledišta*. Jastrebarsko: Naklada Slap.
- Guilford, J. P. (1968). *Osnovi psihološke i pedagoške statistike*. Beograd: Savremena administracija.
- Matlock-Hetzel, S. (1997). *Basic Concepts in Item and Test Analysis*. Texas: A&M University - <http://www.ericae.net/ft/tamu/Espy.htm> (10.05.2007.)
- Norusis, M. J. (1988). *SPSS/PC+ Advanced Statistics V2.0 for IBM PC/XT/AT and PS/2* (Chapter 2: Identifying Dimensions of Communities: Factor Analysis, B-41 to B-69). Chicago:SPSS Inc.
- Nunnally, J. C., Bernstein, I. H. (1994). *Psychometric Theory*. New York: McGraw-Hill.
- Petz i sur. (1992). *Psihologijski rječnik*. Zagreb: Prosvjeta.
- Rathus (2000). *Temelji psihologije*. Jastrebarsko: Naklada Slap.
- Raven, J., Raven, J. C., Court, J. H. (1998). *Priručnik za Ravenove progresivne matrice i ljestvice rječnika*. Jastrebarsko: Naklada Slap.
- Sternberg, R. J. (2005). *Kognitivna psihologija*. Jastrebarsko: Naklada Slap.
- Vasta, R., Haith, M. M., Miller, S. A. (1998). *Dječja psihologija*. Jastrebarsko: Naklada Slap.
- Zarevski, P. (2000). *Struktura i priroda inteligencije*. Jastrebarsko: Naklada Slap.

7. PRILOG

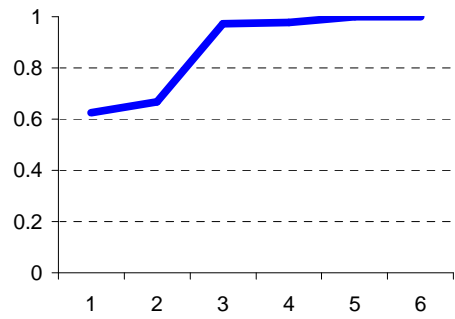
z1



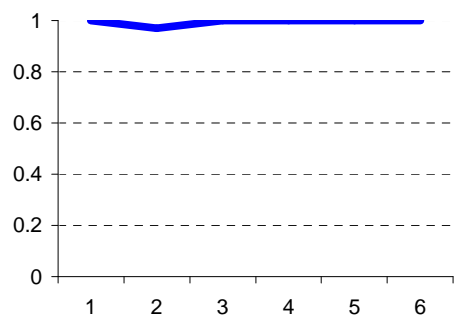
z2



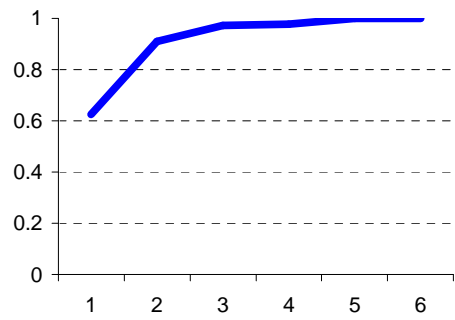
z3



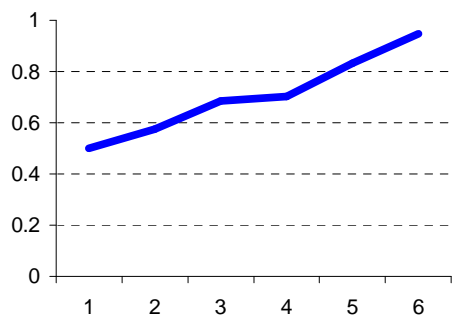
z4



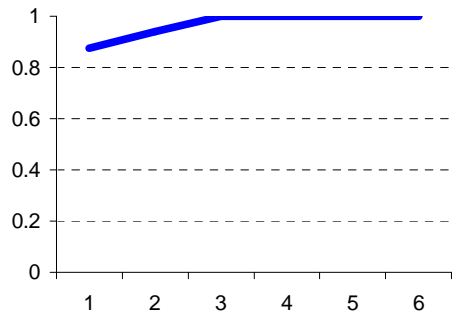
z5



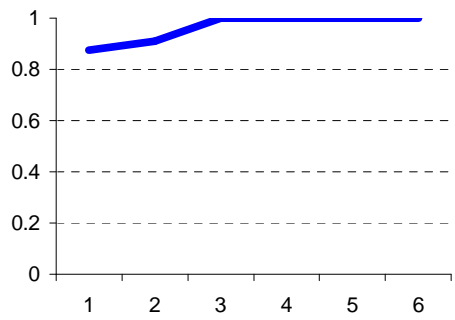
z6



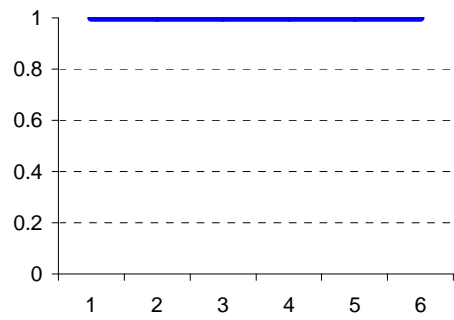
z7



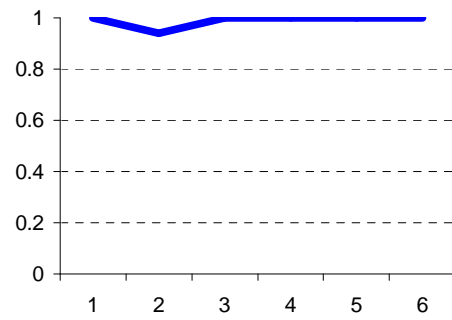
z8



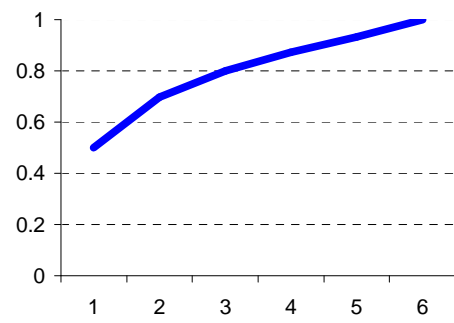
z9



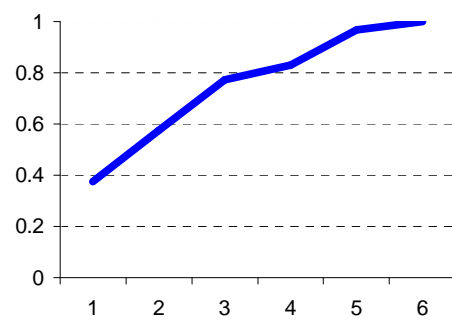
z10



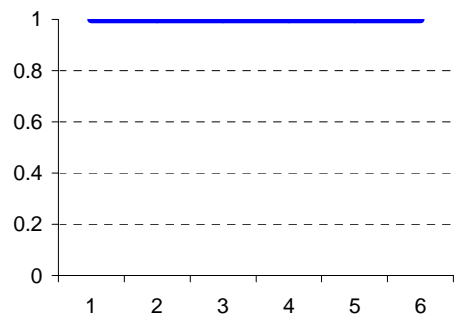
z11



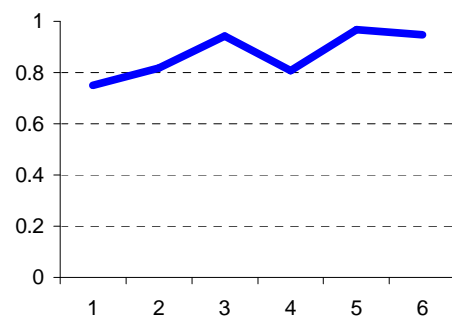
z12



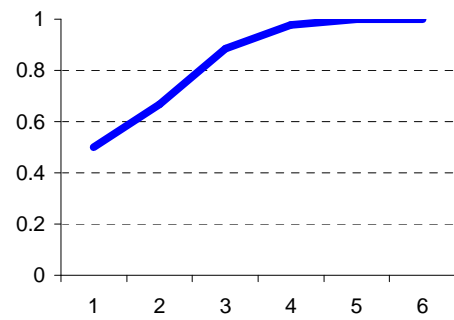
z13



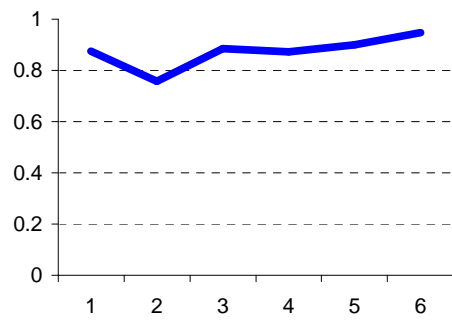
z14



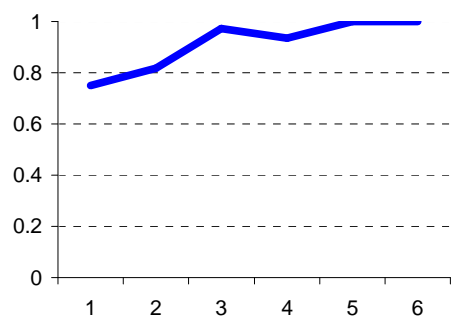
z15



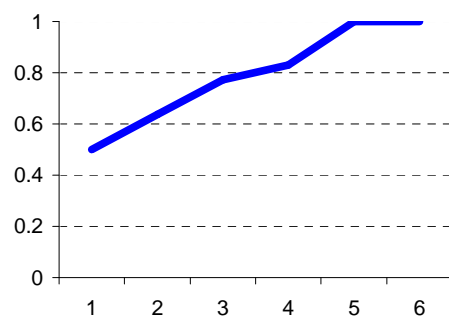
z16



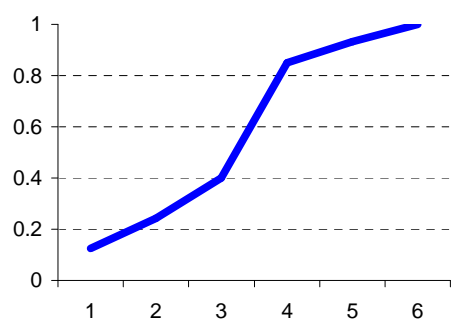
z17



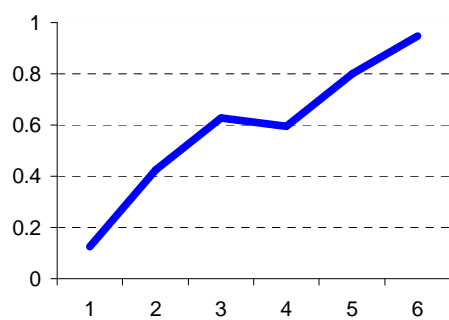
z18



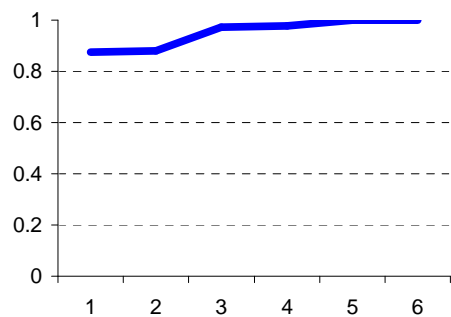
z19



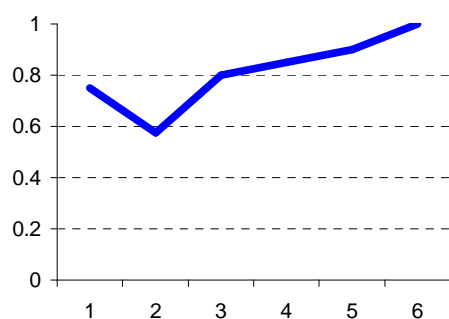
z20



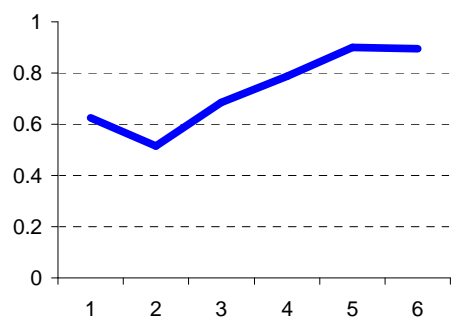
z21



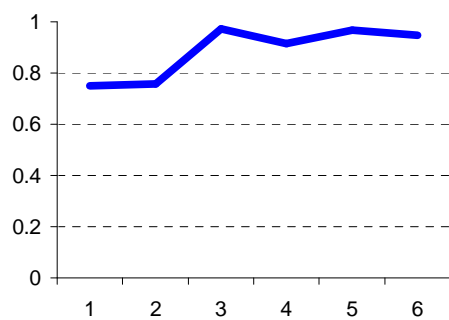
z22



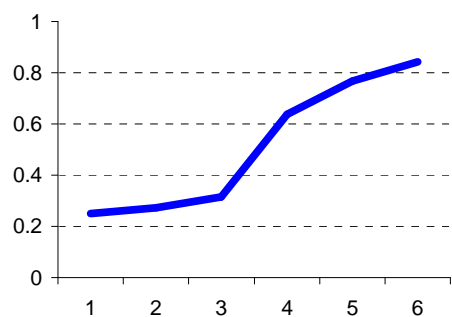
z23



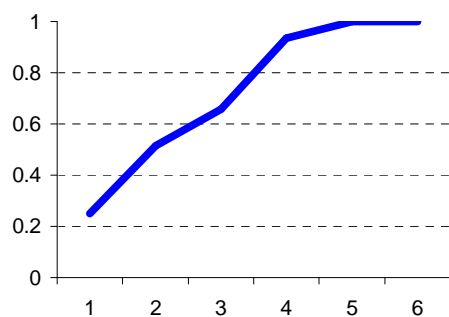
z24



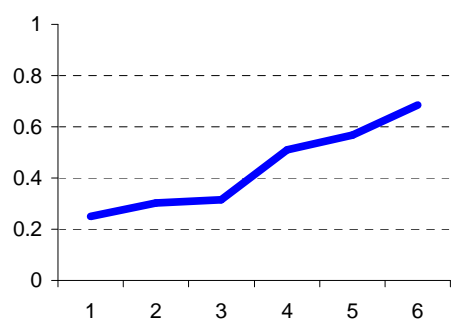
z25



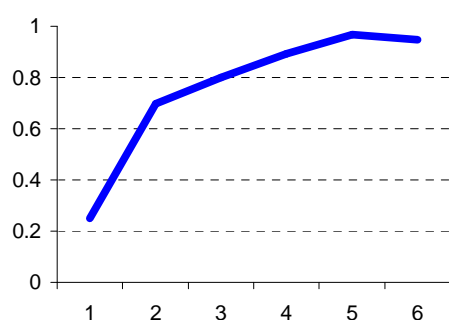
z26



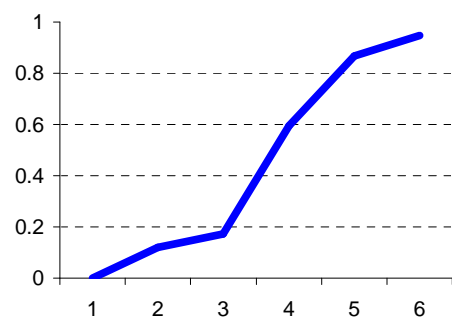
z27



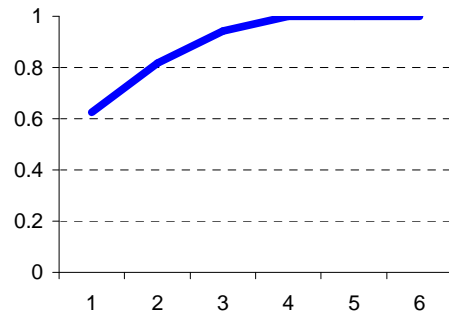
z28



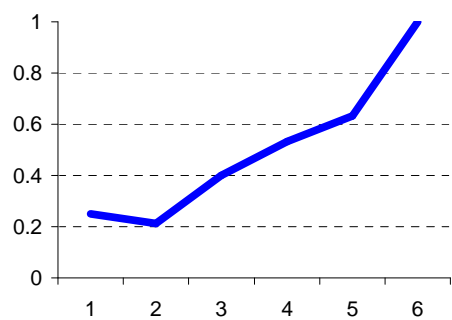
z29



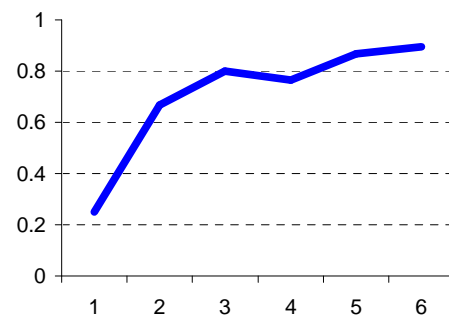
z30



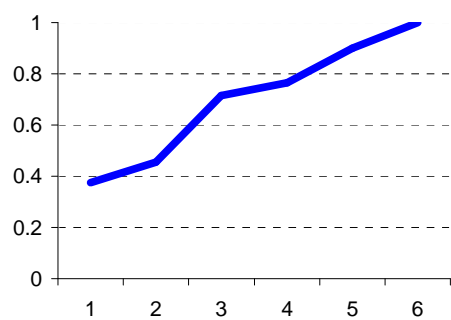
z31



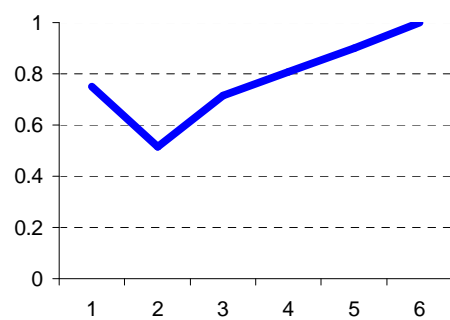
z32



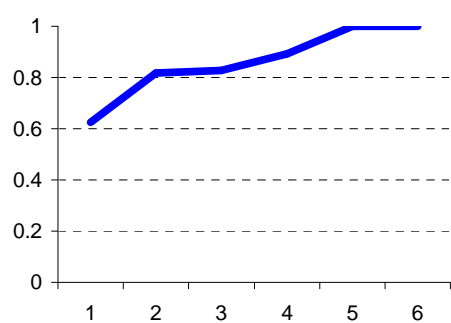
z33



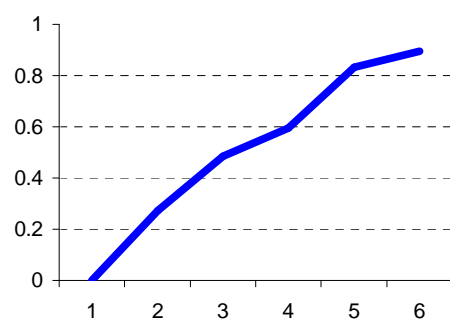
z34



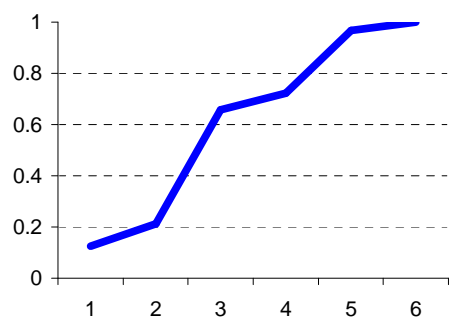
z35



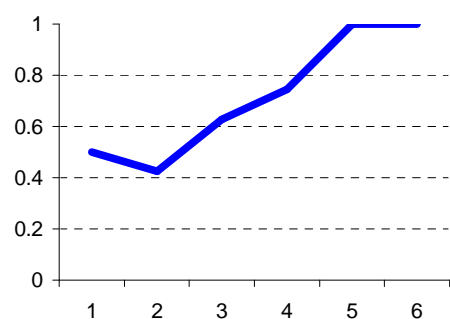
z36



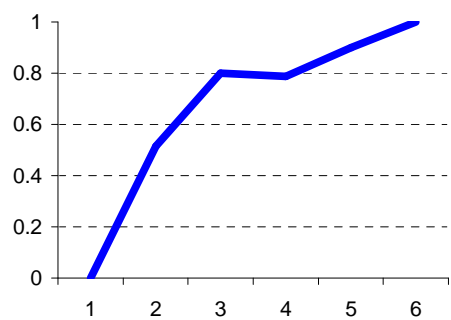
z37



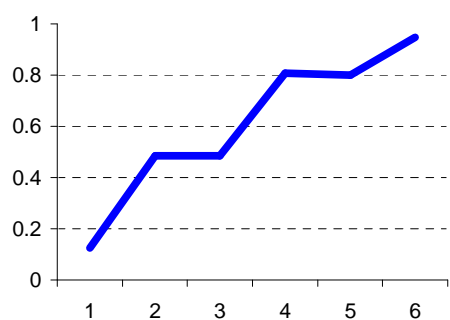
z38



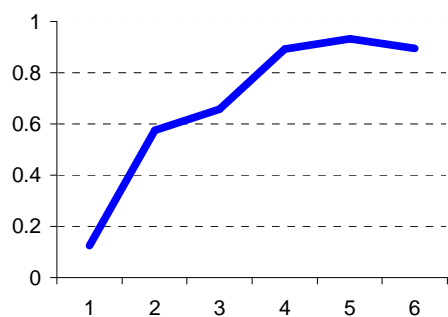
z39



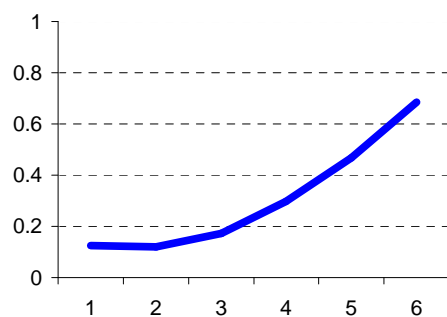
z40



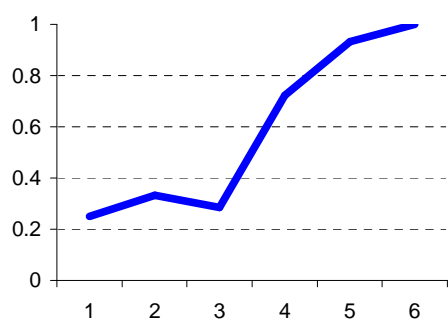
z41



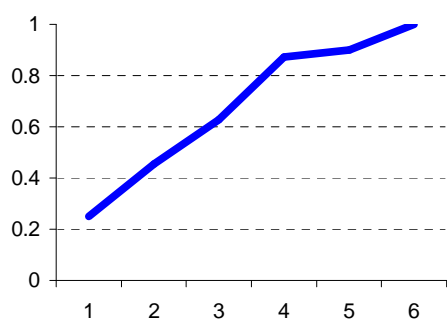
z42



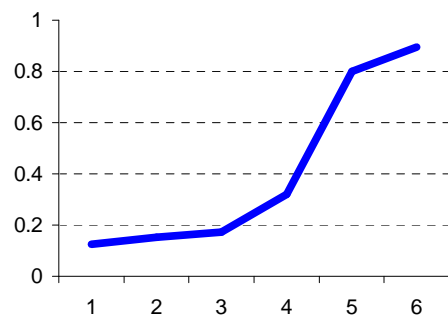
z43



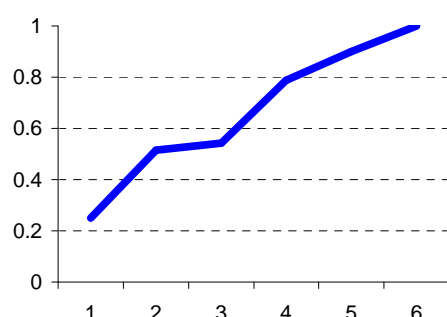
z44



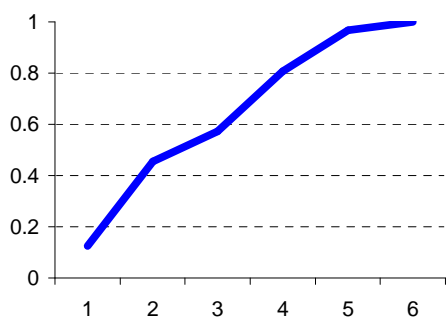
z45



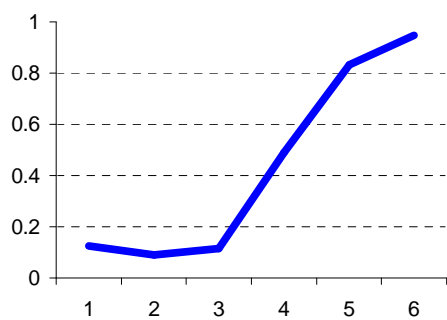
z46



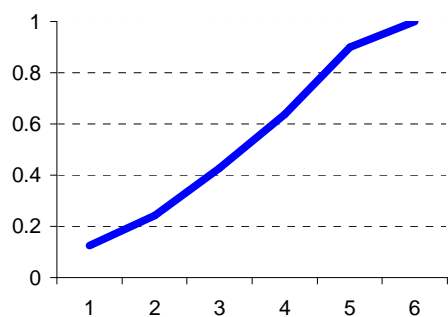
z47



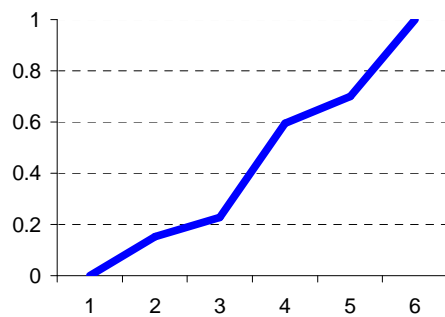
z48



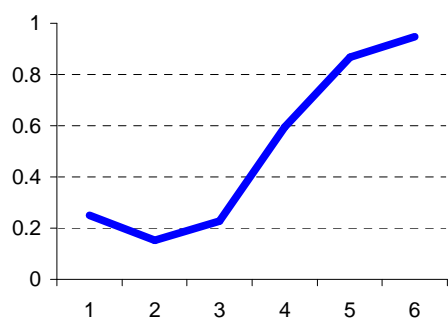
z49



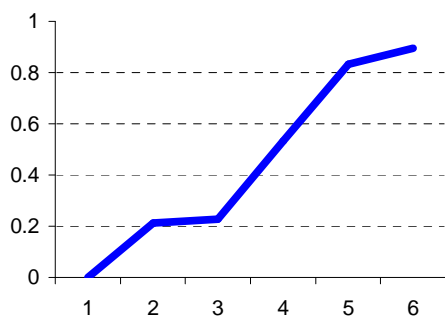
z50



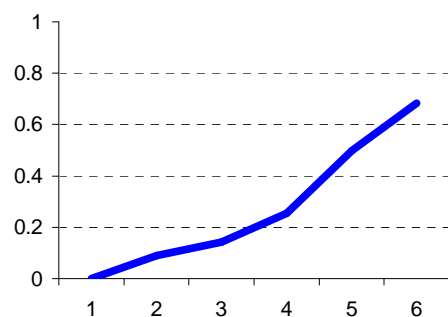
z51



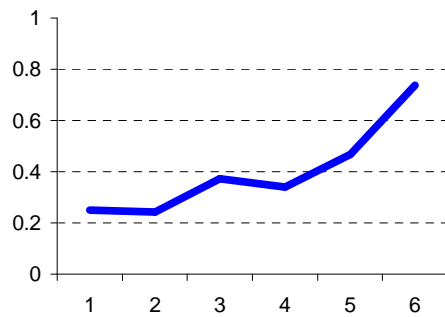
z52



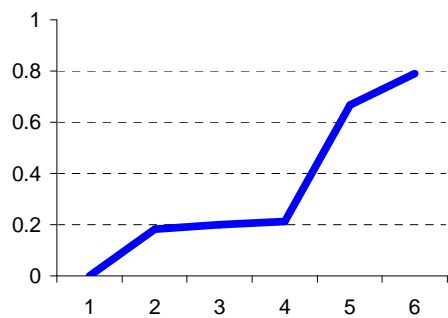
z53



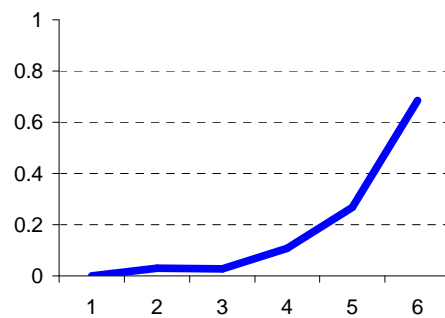
z54

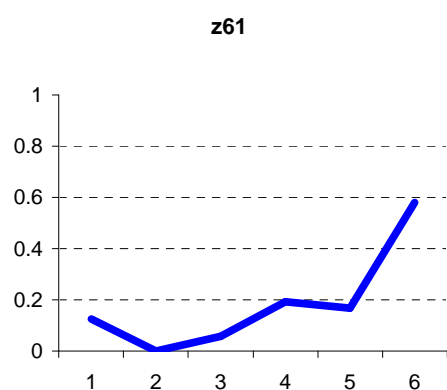
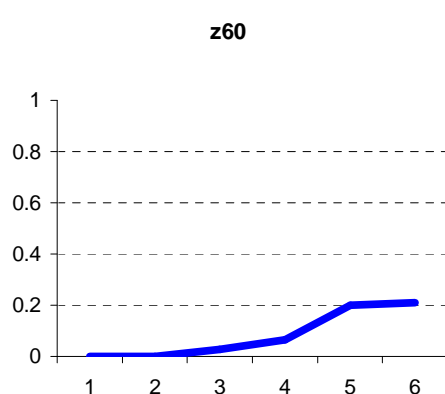
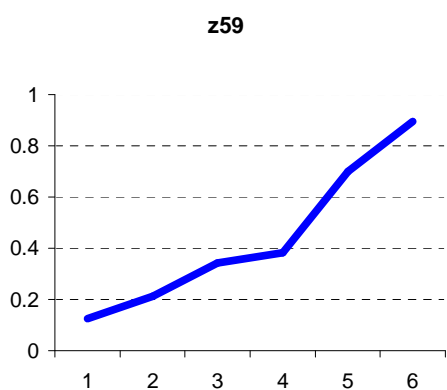
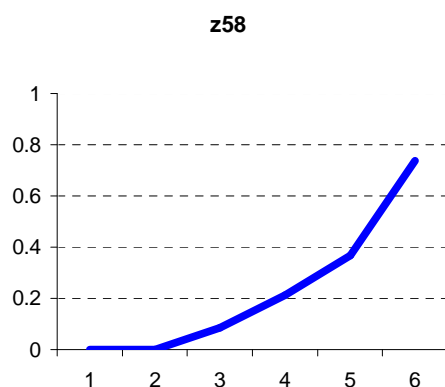
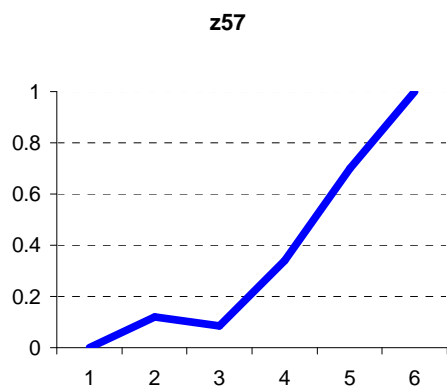


z55



z56





LEGENDA:

razred	ukupni rezultat
1	20-26
2	27-33
3	34-40
4	41-47
5	48-54
6	55-61

Slika 5 . Empirijske karakteristične krivulje 61 zadatka Testa S
 (apscisa: ukupni rezultat na Testu S kategoriziran u 6 razreda;
 ordinata: vjerojatnost točnog odgovora na zadatak - indeks lakoće)